View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by ZENODO

The European Open Science Cloud Pilot Phase

Executive Summary

This document outlines an approach for a limited-scale European Open Science Cloud (EOSC) pilot to be implemented via INFRADEV-04-2016. The objective for this EOSC pilot is to provide confidence that a future full-scale EOSC will promote open science, be fit for purpose, inclusive, sustainable and fully inter-operable with existing systems and installed capacity. The proposed approach is the result of a wide consultation and addresses the recommendations of the European Commission's High Level Expert Group (HLEG) on EOSC established in the latter part of 2015.

The paper poses four interconnected questions about EOSC that need to be answered:

- What is needed?
- How can it be built?
- How will it be managed?
- How can it be funded?

EOSC should deliver trusted access to services & systems in a federated environment by leveraging existing services, across Member States and disciplinary, social and geographical borders, where data complies with the FORCE11 "FAIR" principles (Findable, Accessible, Interoperable, Re-usable).

The approach takes into account the requirements of a number of key stakeholder groups:

- research infrastructure providers
- their public and private sector research communities including the *long tail of science*
- private sector users of scientific results
- e-infrastructure providers, including commercial cloud-service providers and relevant standards bodies
- international, national and regional funding agencies

A set of overarching issues have been identified including the services to be provided and their essential features that will promote open science; a legal framework that must balance the interests of users and service providers while respecting relevant legislation and encouraging innovation; the funding and provisioning models under which services are procured, operated and used; the principles of consensual governance, including compliance with the legal and ethical requirements, that those stakeholders are being asked to buy into.

A key finding is that to achieve the goals outlined by the HLEG on EOSC report will require a significant increase in funding compared to that invested by the stakeholders today. Improving cost-effectiveness through new technology, better governance and innovative funding models will certainly help but will not offset all the increased costs of making a wider range of interoperable services available to far more users.

The pilot EOSC relies on Research Infrastructures and e-Infrastructures that are supported by member states and the EC working closely together as trusted partners. It is recognised that a 'one size fits all' solution is not appropriate and so selected representatives for specific classes of applications from a range of scientific disciplines will act as flagships and verify the suitability of the pilot.

The pilot EOSC platform will offer an aggregation of production quality services drawn from both public and private sources and presented through a common service catalogue. The services must conform to a minimum set of technical, legal and security requirements that ensure their suitability for inclusion. The platform will support a wide range of use cases initially drawn from those user communities and research infrastructures that formally engage with the pilot starting with an initial set of applications from Bioinformatics, Environmental Research (including climate which also makes use of HPC services provided by PRACE), High Energy Physics, Research with Photon Science and the long tail of Science. The 50 Research Infrastructures that feature in the ESFRI 2016 roadmap are all potential stakeholders in the future full-scale EOSC.

The EOSC funding model is complex because it has to balance the use of public and private sector funding, a mix of capital and operational budgets, involvement of commercial service providers as well as public-sector e-infrastructure operators and a broad spectrum of stakeholders. There are, in fact, separate models for capital investment, operational charges and charging structures for scholarly publishing. Funding schemes that vary from country to country and between communities need to be carefully assessed and reflected in the selected models. Clear commitments and incentives are identified to be assessed annually to encourage the engagement of each stakeholder group.

A multi-stakeholder governance model will be established to collaborate with and not replace the existing governance models for research infrastructures, user communities as well as the projects and initiatives that provide services. The consortium for the INFRADEV-04-2016 project should be compact, facilitate the participation of the wider communities in the pilot EOSC, and be composed of partners with good visibility in those communities. The consortium should include organisations leading the flagship applications, coordinating bodies for the engaged public e-Infrastructures as well as open science and open access policy makers.

A limited-scale EOSC pilot can be implemented within the €10m budget of the INFRADEV-04-2016 call by building on existing developments and e-infrastructures to engage stakeholders to determine if the EOSC can deliver on its expectations.

The proposed approach will install a service-oriented platform which is fundamental for European competitiveness through developed funding models & tangible operational plans for a range of services that organisations and end-users want and need. The output will be a tested, limited-scale deployment governed by multiple stakeholders with a clearly-defined legal framework supporting innovative funding models. A roadmap will be produced describing what investments are necessary to grow the pilot so it serves more research infrastructures and data providers, a broader user base from the public and private sectors and a dynamic service platform.

Prepared by Bob Jones (IT department) on behalf CERN

Copyright CERN © 2016.
This document is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Contents

Introduction

The objective of this paper is to propose an approach for the active participation of multiple stakeholders in the establishment of the European Open Science Cloud (EOSC) that builds on their strengths and current engagements in the open science landscape.

The paper poses four interconnected questions about EOSC that need to be answered:

- What is needed?
- How can it be built?
- How will it be managed?
- How can it be funded?

A limited-scale pilot EOSC that can implemented via the INFRADEV-04-2016 EC funding call (deadline 22 June 2016) is proposed as the vehicle to provide answers to these questions.

The following bodies have been consulted in the preparation of this paper:

- EIROforum[1] (CERN, EMBL, ESA, ESO, ESRF, EUROfusion, European XFEL and ILL)
- EU-T0[2] (CERN, CIEMAT, DESY, IFAE, INAF, IN2P3, INFN, IRFU, KIT, STFC and SURFSARA)
- EGI.eu (EGI), CSC (EUDAT), GÉANT

Background on the European Open Science Cloud (EOSC)

The Research Infrastructures component of the HORIZON 2020 Work Programme for 2016-2017 addresses the requirement for a sustainable approach to exploitation of the data produced and/or collected by research infrastructures (RIs). In particular, the action INFRADEV-04-2016 foresees the evolution of existing e-Infrastructures into a 'European Open Science Cloud' (EOSC). The stated aim is to launch a pilot action that should demonstrate how a cloud infrastructure can make scientific data and data-analysis services more widely available enabling greater data sharing and re-use by European researchers. EOSC should deliver a trusted access to services & systems in a federated environment by leveraging existing services, across Member States and disciplinary, social and geographical borders, where data complies with the FORCE11 "FAIR" principles

- Findable
- Accessible
- Interoperable
- Re-usable.

Note that a 'Research and Innovation Action' under the INFRA programme has to address three things: networking with stakeholders, enabling access to data services and improving access through collaborative research.

[1] http://www.eiroforum.org/
[2] http://www.eu-t0.eu/

The EOSC High Level Expert Group

The European Commission established the High Level Expert Group (HLEG) 'European Open Science Cloud' in the latter part of 2015 with a mandate to advise on the scientific services to be provided on the cloud and on its governance structure while recognising that *"a fresh look is required into how to address the need of a data infrastructure that builds on what exists, caters for the whole scientific community and provides the governance and services that are today missing"*.

Many position papers about the EOSC were submitted to the HLEG. EGI, together with other leading European initiatives EUDAT, LIBER, OpenAIRE and GÉANT, have shared their joint vision for the European Open Science Cloud for Research[3] with eight elements of success for a concrete contribution to the Digital Single Market:

1. Open in design, participation and use
2. Publicly funded & governed with the 'commons approach'
3. Research-centric with an agile co-design with researchers and research communities
4. Comprehensive in terms of universality and inclusiveness of all disciplines
5. Diverse & distributed empowering network effects
6. Interoperable with common standards for resources and services
7. Service-oriented as well as protocol-centric
8. Social connecting diverse communities

The EIROforum IT working group published a statement of intent and position paper[4] on the European Open Science Cloud. The paper describes a vision of a cloud-based infrastructure offering freedom of choice to the operators of data-intensive research infrastructures and innovation in provisioning of services for their user communities. To achieve this, EOSC must take a bottom-up approach, starting where there is the most commonality, while making higher-level and thematic services (including those identified and provisioned by each research domain) accessible within a common framework.

Objectives

Our objective for this pilot is to provide confidence that a future full-scale EOSC will promote open science, be fit for purpose, inclusive, sustainable and fully inter-operable with existing systems and installed capacity. The commitment of stakeholders to EOSC and Open Science in general requires:

- A sustainable long-term funding model which recognises the importance of stimulating innovation
- Understanding the focus, scale and distribution of investments needed by the stakeholders to support a wide range of sustainable science activities through a shared platform
- Recognition that the full value of scientific data will be realised by serving the interest of industries capable of leveraging that data.

[3] https://documents.egi.eu/public/RetrieveFile?docid=2637&version=1&filename=OSC_Position_Paper.pdf
[4] https://ec.europa.eu/research/openscience/pdf/eosc-workshop-11-2015/eiroforum_27_november_2015.pdf

To demonstrate we can achieve these objectives, we propose that INFRADEV-04-2016 enables a limited scale deployment of a service platform. The services will be managed via a well-defined lifecycle and accessible through a service catalogue by users from both the public and private sectors. The pilot will aggregate existing services from a variety of sources and projects to provide a platform on which further innovative services (funded through other streams and actions) can be deployed and make them seamlessly accessible to end-users with the intention of demonstrating the feasibility of the EOSC. Success of the pilot will be measured in terms of the engagement of stakeholders:

- Number of research infrastructures that use the EOSC to provide access to the data they produce
- Range and number of services that are offered to end-users
- Scale and diversity of user communities that make use of the services
- An increase in the adoption of open science policies by the stakeholders
- An increase in the sharing and re-use of research data
- Ability of the funding models to support EOSC beyond the pilot phase

The stakeholder engagement will be measured via metrics and Key Performance Indicators (KPIs) that will gather information and data from reliable sources and make use of tools available via contributing infrastructures and projects. For example, the THOR project provides a dashboard[5] which monitors the evolution of persistent identifier (DOI) interoperability which can provide a measure for the sharing and re-use of data of research data.

Analysis

The report from the EOSC High Level Expert Group as well as the input the group received has been analysed together with material presented and discussions held at a number of events around Europe where EOSC has been on the agenda. The implications of this analysis and how the results can be implemented in the pilot EOSC are presented in the sections below. A key finding is that to achieve the goals outlined by the report then EOSC will require a significant increase in funding compared to that invested by the stakeholders today. The HLEG report proposes to make the EOSC available to all European researchers. There were 1.73 million researchers (in full-time equivalents (FTE)) employed in the EU-28 in 2013[6]. In comparison, EGI, Europe's largest publicly funded grid/cloud e-infrastructure, had 57'000 estimated users in 2016[7]. Expanding existing e-infrastructures to be capable of handling more than an order of magnitude more users will require significant investments. Improving cost-effectiveness through new technology, better governance and innovative funding models will certainly help but will not offset all the increased costs of making a wider range of interoperable services available to far more users.

[5] http://dashboard.project-thor.eu

[6] Eurostat statistics on R&D personnel http://ec.europa.eu/eurostat/statistics-explained/index.php/R_%26_D_personnel

[7] T. Ferrari, EGI figures, users registered in VOs or accessing the infrastructure via portals, March 2016

The e-infrastructures landscape, transversal to all research domains, has been elaborated by the e-Infrastructure Reflection Group (e-IRG) as part of the 2016 edition of the ESFRI Roadmap[8]. The 50 ESFRI Research Infrastructures that feature in the roadmap as projects and landmarks are all potential stakeholders in the future EOSC.

Approach

The approach presented here takes into account the requirements of a number of key stakeholder groups:

- research infrastructure providers
- their public and private sector research communities including the *long tail of science*
- private sector users of scientific results such as customers for data from EMBL-EBI and ESA
- e-infrastructure providers, including commercial cloud-service providers and relevant standards bodies
- international, national and regional funding agencies

From these, we have identified five overarching sets of issues:

- the services to be provided
- the essential features of those services that will promote open science, including standards
- a legal framework that must balance the interests of users and service providers while respecting relevant legislation and encouraging innovation
- the funding and provisioning models under which services are procured, operated and used
- the principles of consensual governance, including compliance with the legal and ethical requirements, that those stakeholders are being asked to buy into

The scope of the EOSC is well beyond what can be achieved with a limited-scale pilot to be funded via the INFRADEV-04-2016 call and so the approach relies on working closely with Research Infrastructures and e-Infrastructures that are supported by member states and the EC. It is also recognised that a 'one size fits all' solution is not appropriate and so it is proposed to carefully select representatives for specific classes of applications from a range of scientific disciplines that can act as flagships and verify the suitability of the pilot.

[8] 2016 ESFRI Roadmap http://ec.europa.eu/research/infrastructures

Services

The pilot EOSC platform will offer an aggregation of production quality services gathered from a variety of sources and presented through a service catalogue in a manner similar to the AppStores found on Smartphones. As with AppStores, the services available via the pilot EOSC platform must conform to a minimum set of technical, legal and security requirements that ensure their suitability for inclusion. The services will be drawn from both public and private sources and will be offered for use, as with AppStores, either for a fee or free of charge. The level of usage of all services will be monitored by the platform. Registered users of the platform can apply for grants corresponding to service credits which can then be used to pay for any services in the service catalogue that charge a fee. Service providers set their own prices which must be visible in the service catalogue. Service providers will be paid according to the credits consumed. Service credits will be allocated to end-users from the EC funds attributed to the project. Public service providers that charge a fee for their services will be able to combine multiple funding streams while avoiding double financing. All services will be made available under equitable terms and conditions based on financial and legal policies established with insight into scientific and societal needs and technological possibilities. Such policies will ensure service interoperability and encourage open science, respecting European legislation and the intellectual property of users and data providers.

In this pilot implementation, a core set of interoperable services will be included in the **service catalogue**:

- Data management (including data storage and transfer)
- Compute (including HPC)
- High-level tools that build on the underlying compute/data infrastructure
- 'Human' services: consultancy, s/w development, training etc.

The consultancy and training services are essential if the pilot EOSC is to broaden its user base. Support for uploading, porting and execution of user application codes and datasets will be an essential feature of the platform.

A common **metadata catalogue** will include all datasets accessible via the data services together with a search/discovery facility.

The following initiatives and projects will be invited to contribute services that will initially populate the service and metadata catalogues:

- Data: EGI, EUDAT, Helix Nebula, OpenAIRE
- Compute: EGI, Fortissimo[9], Helix Nebula, Ubercloud[10]
- High-level tools: CloudFlow[11], CloudSME[12], Indigo-DataCloud, etc.
- Human expertise: e-infrastructure competence centres, centres of excellence, EIT Digital and relevant projects such as EDISON which aims to establish the data scientist as a profession, etc.

[9] https://www.fortissimo-marketplace.com/infopage/
[10] http://www.theubercloud.com
[11] http://www.eu-cloudflow.eu/
[12] http://cloudsme.eu/

The introduction of a new service into the service catalogue, including thematic and community-specific services not mentioned in the list above, will be orchestrated to ensure interoperability with other services. The services must offer a secure user Authorisation and Authentication mechanism and be accessible via the GEANT network. A centralised help desk linking together the help desks of the service providers, will also be available for users. A network of Computer Security Incident Response Teams (CSIRTs) will also need to include all service providers.

Essential features and standards

EOSC must address four criteria for these services in order to achieve reproducible inter-disciplinary research and bridge academia and industry.

- The entire data lifecycle has to be actively managed in order to make data findable and accessible.
- The approach to inter-operability must be vendor neutral and globally supported.
- The mechanism for making data secure and re-usable over time must be sustainable.
- The implementation must adhere to relevant standards, notably for security and interoperability, such as those identified by ETSI Cloud Standards Coordination[13].

These essential features and standards will be published as a set of requirements for conformant services. Only services that conform to the requirements will be eligible for inclusion in the service catalogue. The requirements and set of relevant standards for conformant services and their operators will be kept to a minimum to avoid stifling innovation.

Use Cases

The EOSC platform will support a wide range of use cases initially drawn from those user communities and research infrastructures that formally engage with the pilot.
It is proposed to start with an initial set of flagship applications from the existing ESFRI research Infrastructure clusters and the long tail of Science. To broaden further the user base, the project will encourage end-users to propose their own use-cases. These applications will be evaluated to assess their research merit, potential impact and technical compatibility with the platform. Successful applicants will be awarded grants corresponding to service credits to pay for services available in the service catalogue (including consultancy services) in collaboration with the future EINFRA-12-2017 projects (separate but interrelated projects). It is essential that the platform can accommodate lead users from industry (including SMEs) and the public sector to act as a testbed for innovative cloud technology solutions that can develop an ecosystem to strengthen the cloud industry in Europe[14].

[13] http://csc.etsi.org/
[14] https://ec.europa.eu/digital-single-market/en/news/communication-european-cloud-initiative-building-competitive-data-and-knowledge-economy-europe

Legal framework

The EOSC will need to operate within the context of a legal framework that protects the interest of all the stakeholders and ensures conformity with European and national legislation. The EOSC will be part of an on-demand economy where legislation is often inchoate and so its legal framework will have to respond to statutory and non-statutory requirements. This legal framework will take into account Europe's General Data Protection Regulations (GDPR) together with a clear position on jurisdiction, status of privileges and immunities as well as provisions for dispute resolution. Non-statutory constraints are essentially components of the legal relationship between a researcher using EOSC services and the providers of those services. Whereas, for example, privacy is a regulatory issue, preserving trust will be a contractual requirement addressed by a combination of policy, contract and codes of practice. Conformance with the legal framework will be part of the requirements for services to be included in the service catalogue.

Funding models

The EOSC funding model is complex because it has to balance the use of public and private sector funding, a mix of capital and operational budgets, involvement of commercial service providers as well as public-sector e-infrastructure operators and a broad spectrum of stakeholders. There are, in fact, separate models for capital investment, operational charges and charging structures for scholarly publishing.

Development costs can be supported via streams such as institutional/national resources, the Horizon 2020 E-infrastructure work programme, pre-commercial procurement (PCP), public procurement of innovation (PPI), European Structural and Investment Funds (ESIF), European Fund for Strategic Investment (EFSI) as well as investment by commercial cloud service providers, etc.

With regard to operational costs, the underlying premise is that they will be reimbursed according to metered usage while the services will be accessible to end-users as free at the point of use. This aspect will be developed in collaboration with projects funded via the upcoming EINFRA-12-2017 call.

Funding schemes vary from country to country and between different communities, for example sometimes computing costs are included in grants, sometimes not. The breadth of these funding schemes needs to be carefully assessed reflected in the selected models.

It is important that communities beyond the public research sector use the services and so contribute to their funding. Ensuring the EOSC remains as close as possible to mainstream IT trends will facilitate wider usage and spread the development and operational costs thereby improving sustainability.

Stakeholders will be provided with regular, accurate and timely information about the status of the platform and individual services including quality levels, market penetration by sector and user community, etc. The pilot EOSC must quantify the current needs of the research community and their likely growth in order to prepare a roadmap of investments necessary to expand the pilot to full-scale.

Stakeholder commitments and incentives

Clear commitments and incentives are negotiated and re-assessed annually to encourage the engagement of each stakeholder group:

- research infrastructure providers
    - *commitment* – make data available to approved users via the pilot EOSC and promote participation and its use to user communities
    - *incentive* – greater re-use and exploitation of research data leading to improved ease of use and increased impact of the Research Infrastructures
- research communities including the *long tail of science*
    - *commitment* – participation in the pilot EOSC, monitor and report on services consumed and provide feedback
    - *incentive* – service credit allocation for easy access to a wider range of datasets and interoperable services covering the full research life-cycle, opportunity to deploy their own test-cases
- private sector users and resellers of scientific results
    - *commitment* – participate in the pilot EOSC, using services on a pay-per-use basis, report on services consumed and provide feedback
    - *incentive* – easy access to a wider range of datasets and interoperable services covering the full research life-cycle and a platform on which they can offer their own services
- publicly funded e-infrastructures
    - *commitment* – offer conformant services to a wide user base, report on services consumed and provide feedback
    - *incentive* – be part of a vibrant service platform that will promote their services to a wider user base and be eligible for reimbursement of operational costs
- commercial cloud-service providers
    - *commitment* – offer conformant services via the service catalogue, report on services consumed and provide feedback
    - *incentive* – additional revenue streams generated through the service platform
- international, national and regional funding agencies
    - *commitment* – promote participation in EOSC via their funding activities and evaluate metered usage as a basis for funding the operational costs of e-infrastructure services
    - *incentive* – a common platform to share the cost of developing and operating e-infrastructure serving a wider user base and engaging both the public and private sectors

The service platform will support funding models that allow the data market to enter a faster growth trajectory. The supply-demand dynamics will change from technology-push to demand pull. This is a classic virtuous cycle mechanism where network effects multiply the benefits for users in their interactions and makes it easier to consolidate standards and interoperability, reducing further the barriers to adoption.

Governance

The EOSC pilot Governance Model will be composed of 3 elements:

- **Principles of Governance** that express how the 'organisational structure' and 'governance processes' should be set up;
- **Organisational Structure** that define the different levels of responsibility, the roles of the governance bodies, who participates and who influences them;
- **Governance Processes** that formalise the set of activities (including their inputs and outputs) performed by each governance body (i.e. what they do and how they interact).

The EOSC pilot governance model will enshrine a series of widely accepted principles:

- Ensure alignment with the Digital Single Market, foster coherence, equitability and inclusiveness
- Enable integration of existing e-Infrastructures with commercial cloud services effectively and efficiently
- Ensure participation of all stakeholders and fair balance of their needs and interests
- Ensure transparency, openness and responsiveness
- Ensure value for money and fair incentives and returns
- Continuously manage legal and ethical compliance and other risks
- Ensure accountability and responsibility of stakeholders and decision makers
- Manage the identity and brand of a European Open Science Cloud and ensure sustainable innovation and growth.

The pilot EOSC organisational structure will collaborate with and not replace the existing governance structures for research infrastructures, user communities as well as the projects and initiatives that provide services. This pilot EOSC organisational structure will be separate from that of the INFRADEV-04-2016 project itself which will be defined by a DESCA based H2020 consortium agreement.

Inspiration for the organisational structure of the multi-stakeholder governance model can be sought in the Internet Corporation for Assigned Names and Numbers (ICANN) post-transition from the stewardship of the U.S. Government. Given the rapid evolution of the technology and domain, it is essential to maintain a practical approach such as that embodied by the Internet Engineer Task Force (IETF) where “rough consensus and running code” is the rule. Useful input can also be provided by the IT governance model for the Connecting Europe Facility (CEF)[15].

Interaction with member states, who will be the major public funders of the EOSC, will be handled via a number of existing structures including e-IRG for e-infrastructure aspects and forums such as the Belmont Forum[16] for environment research for domain specific aspects. Interaction with research infrastructures will be via ESFRI (notably the *Working Group on investment strategies in e-infrastructures* and its successors), EIROforum and the EC funded ESFRI cluster projects such as ASTERICS, CORBEL, EMBRIC, ENVRI+, PARTHENOS, SERISS and SINE2020. Engagement with existing e-infrastructures will be via European-level projects including EGI-Engage, GEANT GN4 Phase 1, EUDAT2020, OpenAIRE2020 and Indigo-DataClouds as well as those providing community specific services. It is expected that funding for these projects will continue via the H2020-EINFRA-2016-2017[17] work programme so that the relationship can be assured through to the end of the pilot.

[15] https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/Governance

[16] http://www.belmontforum.org/announcements/2016/belmont-forum-adopts-open-data-principles-environmental-change-research

[17] https://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/calls/h2020-einfra-2016-2017.html

Interaction with private sector service suppliers will be via the European Cloud Partnership and related public-private partnerships such as the Helix Nebula initiative, industry groupings such as EuroCloud as well as sector specific downstream industry groupings including the European Association of Remote Sensing Companies (EARSC) and the contractual Public Private Partnership on Big Data Value (BDVA PPP). The Research Data Alliance (RDA) will provide a forum for interacting with data practitioners from many research disciplines within Europe and globally. A user board will ensure engaged user communities have a strong voice together with mechanisms such as the ability to provide feedback on the services and dedicated online and face to face events to gather feedback from individual users, including the long-tail of science. The proposed High-Level Advisory Group 'Open Science Policy Platform' (OSPP)[18] would be a suitable body to review annually the progress of the pilot EOSC.

To ensure the pilot EOSC can serve global research communities the project will collaborate internationally with funding agencies and structures that are also pursuing a service platform approach (including the NCI Cancer Genomics Cloud Pilots[19] in the USA and Data61[20] from CSIRO in Australia).

Some key roles will also need to be included in EOSC organisational structure. An *innovation officer* responsible for managing the innovation process inside the EOSC will also be required. The tasks include identifying strategies, business opportunities and new technologies and then developing new capabilities and architectures with the stakeholders, including new business models and new structures to serve those opportunities.

A *data protection officer* will be designated with responsibility for making sure that the EOSC follows applicable data regulations. Tasks include informing and advising the stakeholders of their obligations and monitoring the implementation and application of the EOSC's policies. This includes monitoring personal data breaches and the responses to requests from authorities.

INFRADEV-04-2016 consortium

The consortium for the INFRADEV-04-2016 project should be compact, facilitate the participation of the wider communities in the pilot EOSC, and be composed of partners with good visibility in those communities. Establishing the governance model will require senior figures with an international reputation and visibility to be assisted by a secretariat. Representatives participating in the governance structure will be offered a per diem and to have their costs reimbursed. Financial, legal and communication expertise can be contracted from the private sector. The roles of *innovation officer* and *data protection officer* should be funded by the project. The consortium members should ensure the smooth running of the platform including operating the common service and metadata catalogues, billing/invoicing/accounting, on-boarding of new services and users etc.

Implementations for these operational elements already exist, such as those operated by EGI, however dedicated *instances* should be established for the pilot EOSC because they will need to support services from multiple projects and sources while the projects themselves will continue to offer their services via alternative channels. Note that it will not be necessary for the providers of services that are included in the service catalogue to be members of the project consortium.

It is proposed that the consortium includes:

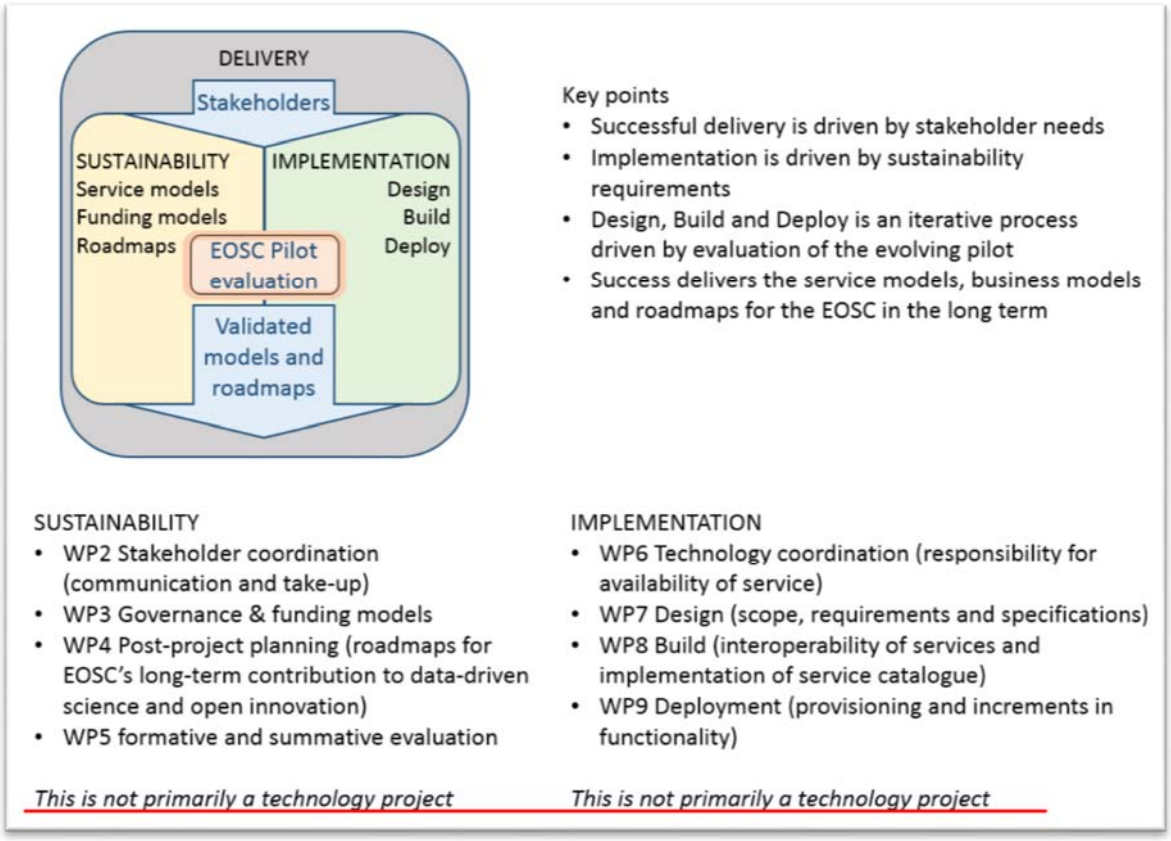
- Organisations leading the flagship applications for engaged research infrastructures including.

[18] http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-policy-platform
[19] https://cbiit.nci.nih.gov/ncip/nci-cancer-genomics-cloud-pilots
[20] http://www.csiro.au/en/Research/D61

- Coordinating bodies for the engaged public e-Infrastructures (EGI, EUDAT, GÉANT, OpenAIRE & PRACE)
- Open science and open access policy makers (e.g. research councils, LIBER).

INFRADEV-04-2016 work plan

This pilot phase for the EOSC should be organised into two steps that collectively address the Implementation and Sustainability of the full-scale EOSC over a period of two years. During the first year of the project, the focus will be on stakeholder engagement, agreeing the governance & funding models and designing the pilot EOSC deployment. During the second, these models will be deployed (in collaboration with EINFRA-12-2017 projects) and evaluated feeding into the roadmap for the full-scale EOSC. The break-down into work packages and their interrelation is depicted in the figure below.



HLEG Recommendations

This approach outlined in this paper is the result of a wide consultation and can actively contribute to the implementation of the HLEG recommendations for the preparatory phase of EOSC, starting with policy:

- P1: Take immediate, affirmative action in close concert with Member States
    - Response: pilot will build on the installed services and capacity of the Member States
- P2: Close discussions about the 'perceived need'
    - Response: This paper presents an implementation plan
- P3: Build on existing capacity and expertise where possible
    - Response: The pilot will build on the installed services and human capacity of the Member States
- P4: Frame the EOSC as supporting Internet based protocols & applications
    - Response: Services must conform with a minimum set of technical, legal and security requirements

This paper reflects conformance with the four HLEG governance recommendations:

- G1: Aim at the lightest possible, internationally effective governance
    - Response: The governance of the EOSC will collaborate with and not replace the existing governance models for research infrastructures, user communities as well as the projects and initiatives that provide services
- G2: Guidance only where guidance is due
    - Response: Guidance will be limited to a minimum set of technical, legal and security requirements
- G3: Define Rules of Engagement for formal participation in the EOSC
    - Response: A clear set of commitments and incentives for each stakeholder group will be re-assessed annually
- G4: Federate the Gems across Member States
    - Response: The pilot EOSC platform will offer an aggregation of production quality services gathered from a variety of sources

This paper supports the seven HLEG Implementation recommendations:

- I1: Turn the HLEG report into an EC approved White Paper to guide EOSC initiative
    - Response: Please distribute the DOI for the White Paper ☺
- I2: Develop, Endorse and implement a Rules of Engagement scheme
    - Response: A set of Rules of Engagement will be negotiated with the stakeholders and published as a deliverable of the Governance activity
- I3: Fund a concentrated effort to locate and develop Data Expertise in Europe

    - Response: The pilot will identify and promote the expansion of human capital starting with the e-infrastructure competence centres, centres of excellence, EIT Digital and relevant projects such as EDISON

- I4: Install a highly innovative guided funding scheme for the preparatory phase
    - Response: The pilot will be supported by funding models including the development and operation of services that draw on multiple funding streams from the public and private sectors and gives users a key role in the decision making process
- I5: Make adequate data stewardship mandatory for all research proposals
    - Response: The pilot EOSC will offer a platform by which data management plans can be implemented
- I6: Install an executive team to deal with international coherence of the EOSC
    - Response: Interaction beyond Europe is part of the Governance activity
- I7: Install an executive team to deal with the preparatory phase of the EOSC
    - Response: The proposed High-Level Advisory Group 'Open Science Policy Platform' (OSPP) would be a suitable body to review annually the progress of the EOSC

The implementation plan also addresses the immediate actions identified through feedback from member states and funding agencies to the HLEG during March 2016:

- II1: Publish the report (final draft available).
    - Response: Please distribute the DOI for the paper ☺
- II2: Develop, Pilot and implement a Rules of Engagement scheme.
    - Response: Incentives and commitments for each group of stakeholders are defined.
- II3: Detail the transition and sustainability model (and pilot it).
    - Response: The pilot will integrate existing infrastructures and propose funding models.
- II4: Train the data experts to bridge between ‘e-INFRA’ and ‘ESFRI’.
    - Response: The pilot EOSC will engage the existing excellence and competence centres which can provide training and consultancy services.
- II5: Assist data stewardship planning and exec. tools for all researchers.
    - Response: The pilot EOSC will offer a platform by which data management plans can be implemented while training and consultancy on data stewardship will also be available.
- II6 Develop the plan for what ‘minimal essential governance’ means in practice.
    - Response: An outline of a governance model is proposed that will be tested during the pilot EOSC.
- II7 Federate Interoperability standards and best practices.

- Response: Interoperability standards will be part of the compliance criteria for services to be included in the service catalogue.

Conclusion

A European Open Science Cloud has the potential to greatly improve the provisioning of IT services for Research Infrastructures to address their big data needs. It can encompass all the phases of the research lifecycle and offer a platform of joint innovation for the public and private sectors. It will offer new approaches to the way IT services are provisioned, organised and funded. The key challenges are integrating frequently changing technologies, managing the complexity and identifying the optimal organisational and financial models. The EOSC is an ambitious undertaking requiring the active engagement of many stakeholders and careful planning of the technical, financial, legal and governance aspects.

The limited-scale pilot EOSC described above can be implemented within the €10m budget of the INFRADEV-04-2016 call by building on existing developments and e-infrastructures to answer such questions and to engage stakeholders to determine if the EOSC can deliver on its expectations.

The approach proposed in this paper will install a seamless service-oriented platform which is fundamental for European competitiveness through developed funding models & tangible operational plans of a series of services that organisations want and need as well as covering market expansion stemming from rapid innovation. The output will be a tested, limited-scale deployment of a service platform governed by multiple stakeholders with a clearly-defined legal framework supporting innovative funding models. A roadmap will be produced describing what investments are necessary to grow the pilot so it serves more research infrastructures and data providers, a broader user base from the public and private sectors and a dynamic service platform.